

*Citation for published version:*

Duggan, GB & Payne, SJ 2001, 'Interleaving reading and acting while following procedural instructions.', *Journal of Experimental Psychology: Applied*, vol. 7, no. 4, pp. 297-307. <https://doi.org/10.1037/1076-898X.7.4.297>

*DOI:*

[10.1037/1076-898X.7.4.297](https://doi.org/10.1037/1076-898X.7.4.297)

*Publication date:*

2001

*Document Version*

Peer reviewed version

[Link to publication](#)

Copyright 2001 by the American Psychological Association. Inc. This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

**University of Bath**

## **Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Running head: INTERLEAVING READING AND ACTING

Interleaving reading and acting while following procedural instructions

Geoffrey B. Duggan and Stephen J. Payne

Cardiff University

Cardiff, United Kingdom

This research was supported by EPSRC Research Studentship Award  
98319806.

Correspondence concerning this article should be addressed to Geoffrey  
Duggan or Stephen Payne, School of Psychology, Cardiff University, P. O. Box 901,  
Cardiff, CF10 3YG, UK. Electronic mail may be sent to [DugganG2@cardiff.ac.uk](mailto:DugganG2@cardiff.ac.uk) or  
[PayneS@cardiff.ac.uk](mailto:PayneS@cardiff.ac.uk).

### Abstract

Memory for an interactive procedure learnt from written instructions is improved if the procedure can be carried out while the instructions are being read. The size of the read-act cycle was manipulated by comparing “chunked” instruction-following, in which 3 or 4 steps are read then enacted, with single-step conditions. In two experiments, enforced chunking improved subsequent unaided performance of the procedure. In Experiment 3 participants were allowed to manage the interleaving of reading and acting. The imposition of a small behavioural cost (a single mouse point-and-click operation) on the switch between instructions and device encouraged more chunking, and better subsequent test performance. We conclude that the interleaving of reading and acting is an important practical concern in the design of interactive procedures, and that more effective chunk-based strategies can quite readily be encouraged.

### Interleaving reading and acting while following procedural instructions

Good procedural instructions for interactive devices must satisfy two criteria. First, they must support performance. Like all procedural instructions they should effectively communicate the procedure they describe, so as to allow users who don't know the procedure to enact it successfully and efficiently. Second, they must support learning. In common with instructions for all procedures that will be used repeatedly, they should facilitate subsequent memory for the procedure, so that it might later be performed without consulting the instructions.

Unfortunately, the nature of human learning is such that these two criteria can sometimes conflict. Schmidt and Bjork (1992) have pointed to a general phenomenon in the relationship between performance during training episodes and later retention of the trained skill. Better performance during training does not always lead to better retention. Training should introduce difficulties for the learner, so as to better model the post-training task. (As Schmidt and Bjork discuss, this general principle of training is strongly related to the transfer-appropriate processing account of several phenomena in the literature on human memory.)

How might procedural instructions be designed so as to follow the Schmidt and Bjork paradigm and provide transfer-appropriate practice opportunities for the learner? Of course, not all manipulations that introduce difficulties during learning are beneficial for the learner. Simply making the instructions unclear is unlikely to be effective, however much this idea may have informed the design of some commercial user manuals. The criterion that quality instructions must communicate the procedure that they describe cannot be ignored.

Some guidance can be obtained from the theoretical literature on text comprehension. This literature introduces a useful distinction between two components of the memory representation derived from reading a text (see Zwaan & Radvansky, 1998; Kintsch, 1998). The textbase represents the propositional information contained in the text itself, and will be the primary contributor to tests like recall of the text. The situation model represents the situation described by the text, it integrates text propositions with inferences and information derived from the reader's background knowledge. The situation model is the primary contributor to tests that go beyond memory for the propositional content of the text. These tests include problem solving and inference making using knowledge derived from the text.

Informed by this distinction, work by McNamara, Kintsch, Songer, and Kintsch (1996), has shown how expository text can be designed to introduce difficulties for readers in exactly the productive manner advocated by the Schmidt and Bjork conception of training. These authors created two versions of target texts, one more coherent than the other (one experiment used a text about traits of mammals, a second used a text about heart disease). Coherence cues were provided by linking clauses with appropriate connectives and by inserting topic headings. The level of readers' background knowledge on the topic of the text was also assessed with a pre-test. After reading a text participants were given tests of the textbase (free recall of the text propositions and specific factual questions about the contents of the text) and tests of the situation model (problem solving based questions, questions requiring inferences from the text, and a concept-sorting task).

McNamara et al (1996) reported that for measures that tested the textbase, the high coherence texts produced better performance. However, for situation model measures, test performance for high knowledge readers was better when they read the

low coherence text. McNamara et al. argued that limiting the coherence of a text forced readers to engage in compensatory processing to infer unstated relations in the text. This compensatory processing supported a deeper understanding of the text, in that the information in the text became more integrated with background knowledge. Thus, for high knowledge readers the texts that were more difficult to read improved the situation model by encouraging more transfer-appropriate processing. Low-knowledge readers were, presumably, unable to achieve the compensatory inferences, and therefore did better with more coherent texts. Because the textbase does not incorporate background knowledge it was not enhanced by any compensatory processing.

The work of Diehl and Mills (1995) further illustrates the relevance of the theory of text comprehension to the design of instruction for interactive procedures. They argue that in the case of procedural instructions the distinction between situation model and textbase maps directly onto a distinction between memory for the procedure (as tested by later task performance) and memory for the instructions themselves.

Texts describing how to complete a task using a device (setting an alarm clock, or constructing a child's toy) were provided. While reading a text participants were required to either perform the task (read and do), or do nothing (read only). (In addition, Diehl and Mills studied some intermediate conditions, such as read and watch experimenter do. These conditions produced intermediate results and are not relevant to the current argument.) The effect of these training methods was then examined by asking participants to recall the text, and then complete the task.

Diehl and Mills reported that the increased exposure to the device in the read and do condition resulted in improved task performance times relative to the read only

condition. However, text recall was better in the read only condition, supporting the conceptual separation of textbase and situation model.

One successful practical approach to the design of instructions for interactive devices is perhaps quite strongly related to this more theoretically oriented work. The concept of a ‘Minimal Manual’ was outlined by Carroll (1990). It sought to minimize the extent to which instructional materials obstruct learning. Crucially, a well-designed Minimal Manual does not necessarily optimize the speed at which users can perform procedures as they read. Carroll’s manuals avoided explicit descriptions that encouraged rapid but mindless rote performance. Instead the emphasis was on active learning whereby learners were encouraged to generate their own solutions to meaningful tasks. This process was facilitated in part by reducing the amount of text provided and including information about error recovery.

Like Carroll, our goal in this article is primarily practical. However, rather than developing a general heuristic framework for instruction, we focus on a particular technique that exploits the idea of transfer-appropriate practice, following the principle of Schmidt and Bjork and the methods of McNamara and colleagues. Like the manipulations of Diehl and Mills, our central interest is not the design of the instructions per se, but rather the way the instructions are read and used. Diehl and Mills’ reported advantage for reading-and-doing over reading alone has no real practical implication, as it is difficult to imagine anyone advocating isolated reading as a preferred method. However, we suggest that the way learners manage the interleaving of reading and doing will affect their later retention, and thus offers an important lever for improving instruction.

Many procedural instructions have a natural step-wise structure, and in these cases it is possible to execute the procedure while reading with minimal load on

memory. Learners can read a single step, then execute it before reading the next step. Such an approach is low on effort (and therefore attractive to the learner), but also low on transfer-appropriate practice and therefore, we predict, poor at encouraging retention. If learners could instead be prompted to read several procedural steps before enacting them, performance would be made more effortful, but learning might benefit. Readers would be encouraged to integrate the information across the chunk of procedural steps, and the increased memory load would provide transfer-appropriate practice.

Our strategy for developing and testing this idea is as follows. First, we report two experiments in which participants are forced into either a step-wise or a chunk-based strategy for interleaving reading and acting. These experiments test our prediction that reading-by-chunks will tax performance during training, but improve learning, in particular retention of the procedure. Next, we report a third experiment which develops a more subtle, indirect manipulation of chunking which we believe holds greater promise of practical application.

## Experiment 1

Experiment 1 manipulated the number of procedural steps participants were forced to read before executing them during training. The development of the textbase and the situation model were then assessed using free recall and task performance respectively.

The main prediction of this experiment, and the one that is most important for practical concerns, is that the increased cognitive effort required to read instructions in



chunks rather than singly at training will improve task performance at test. The secondary predictions concern the textbase. In the studies of McNamara et al. and of Diehl and Mills, manipulations that improved the situation model (task performance at test, in our case) depressed the textbase (recall of instructions). However, we would argue that such a competition between situation model and textbase is not inevitable. Rather, it is critically dependent on the degree to which the textbase and situation model are inferable from each other. Where there is a very close relation between text and situation model, a participant might use memory for whichever has been favoured during training to infer the other at test. With this in mind, we developed two sets of instructions. One set was more elaborate than the other in that it contained more propositions that were not essential to the procedure. These propositions were therefore not easily inferable from the situation model (and vice versa), and should facilitate empirical dissociations between textbase and situation model.

## Method

### Participants

There were 29 female and 3 male participants ranging in age from 19-22 years with a mean of 19.4 years. They completed the experiment in exchange for course credit.

### Stimulus Materials

Each of the experimental tasks required participants to complete a procedure using a computer simulation of a VideoCassette Recorder (VCR). A program that simulated the Toshiba V-727B VideoCassette Recorder was written in Visual Basic 6.

The screen interface presented to the participants had three components. One panel represented a Remote Control. All of the buttons and functions on the simulated Remote Control were in the same position and had the same name as for the actual device, but were operated by mouse clicking. A second panel was used to present the instructions for each task.

The third panel represented a TV screen and was used to display text that would ordinarily appear on the real television screen. The simulation of the television screen differed from an actual screen during operation in three ways. Firstly, any instructions informing participants what buttons to press were removed from the screen in the simulation. Secondly, any information displayed on the actual VCR itself, was displayed along the top of the simulation screen instead. Thirdly, anytime a picture would appear on the real screen, the simulation merely displayed a text message to indicate the operation taking place (i.e., PLAYING, REWINDING etc.).

Four tasks were designed, each pertaining to a different function of the VCR. The TimerProg task entailed using the programming timer to set the VCR to record a program at a particular time of day, and then cancel this action. The VPlus task required the participants to record a programme using the VideoPlus+ function, by entering a particular code and then editing it. The Playback task required participants to carry out the basic “record-now” operation, and then use many of the functions available when playing back a cassette. In the Setup task participants had to manually set the VCR clock, then tune in a channel and store its position. Each task was composed of 14 steps with a distinct instruction to be completed at each step (steps varied from one to several button-clicks, e.g., press “rewind”, or enter a numeric pluscode).

Two sets of instructions were derived from the manual that outlined the general procedure for completing each task. They differed in the level of coherence. This was achieved by adding and deleting linguistic coherence signals, following the procedures outlined in McNamara, Kintsch, Songer and Kintsch (1996), together with some minor re-wordings for the sake of style and consistency. The less coherent instructions are hereafter referred to as “minimal” and their more elaborated counterparts are called “elaborate”. For each task both sets of instructions were broken down into four chunks of three or four steps, and for the elaborate instructions a title was provided for each of these chunks (see Appendix for both sets of instructions from the TimerProg task). The specific parameters for each task, that would ordinarily be provided by the user of a VCR, such as the channel to be recorded or the date on which to record it, were provided on paper in a separate task outline.

The VCR simulation program was run on a PC 5100 Professional. Each time the mouse was used to operate the Remote Control, or check the instructions, the program recorded which button was pressed and time stamped the event. Some of the buttons on the Remote Control did not contain any markings denoting their function. Thus, participants were also provided with a diagram of the Remote Control that identified the function of each of the buttons. This diagram was photocopied from the original manual for the actual VCR.

### Design

The main independent variables were the degree of coherence of the instructions (minimal vs. elaborate) and the chunking of presentation of instructions (chunked vs. single). These were combined in a within-subjects design to produce

four different conditions. Each condition was assigned to one of the four tasks, and each participant performed each task and each condition once. Assignment of conditions to tasks and serial position was done so that across participants each task was assigned equally frequently to each experimental condition, and for every participant elaborate and minimal texts alternated, whereas chunked texts appeared either first and last, or second and third. Further, each one of the four tasks and each of the four experimental conditions appeared equally frequently in each of the four serial positions.

There were two measures of task performance during both the training and the test phase. These were the number of steps on which an error was made, and the time it took to complete the overall task.

### Procedure

Each participant was tested individually in a laboratory. At the start of the experimental session participants were asked how often they used a VCR (never, one or two times a year, month or week, or everyday), and to rate on a scale of 1-5 how competent they considered themselves at using complicated electrical appliances (e.g., video recorders, washing machines, and digital alarm clocks).

Participants were then asked to read through some general instructions describing the video interface and the experimental procedure, before completing a practice task. The practice task was composed of four steps that were not present in any of the experimental tasks. The four experimental tasks were then completed in succession. For each task there was a training phase, followed by a text recall phase, followed by a performance test phase.

Each task outline was presented to participants on a sheet of paper. After reading through the task outline participants were asked to begin the training phase, and were informed that at some point later in the experiment they would be required to complete the same task without instructions. They were not explicitly told that they would have to recall the instructions in writing. The video interface was displayed on the screen with a Start button in the top left hand corner, and an OK button just below it. When the Start button was clicked upon it disappeared, instructions appeared in the Instruction panel, and the program began recording the amount of time elapsed. After participants had read through the displayed instructions and were ready to carry them out, they clicked on the OK button. The instructions and the OK button then disappeared, and the Remote Control now responded to user input. When these instructions had been completed, the OK button reappeared, the Remote Control was temporarily disabled, and the next instructions were presented in the Instruction panel. This process was repeated until participants had completed all 14 steps of the task.

The number of instruction-steps presented at a time in the instruction panel varied according to the condition. In the single conditions, instructions for each step were presented individually, and the step was completed before the instructions for the next single step were presented. In the chunk conditions, three or four steps of instructions appeared at once, all of these were completed, then participants received the next chunk of instructions. The title of each chunk was displayed at the same time as the instruction-steps in both conditions.

If a wrong button (for the current task and procedure) on the Remote Control was selected the computer emitted a beep, and the participant was required to try again to click on the correct button.

Once all 14 steps had been completed, the training phase for that task was over. The task outline was taken away from the participants and they were asked to write down as much as possible of the instructions that had appeared in the instruction box. During recall the computer was switched off but the diagram of the Remote Control was still visible. When the participant had finished writing, the recall protocol was removed and the task outline was handed back.

Participants then had to complete the same task, but this time without the instructions. During this performance test phase if participants failed to complete an instruction step within 30 seconds from the end of the previous step, the experimenter informed participants which buttons needed to be selected to complete the step. They were then allowed to continue unassisted.

The entire train-recall-test cycle was then repeated for the other three tasks. Due to the slight differences in procedures for the chunk and single conditions, before attempting each of these conditions for the first time participants were required to carry out the corresponding procedure during the practice task. This meant each participant completed the practice task on two different occasions.

## Results

A 2 (number of stages: chunk or single) x 2 (coherence: elaborate or minimal) ANOVA with repeated measures on both factors was used for each of the dependent measures. Effect sizes for any differences reported were computed as point biserial correlations.

### Training

For each of the four conditions means were calculated for time spent reading, and time spent executing the instructions, along with errors made during training. The results are given in Table 1.

TABLE 1

The pattern of results shown in Table 1 is straightforward. Less time was spent overall reading the minimal instructions than the elaborate instructions,  $F(1, 31) = 28.00$ ,  $MSE = 747.32$ ,  $p < .001$ . But, per syllable more time was spent reading the minimal instructions than the elaborate instructions,  $F(1,31) = 119.12$ ,  $MSE = .011$ ,  $p < .001$ . When instructions were presented in chunks rather than individual stages participants took longer to execute them,  $F(1,31) = 38.64$ ,  $MSE = 3098.26$ ,  $p < .001$ , and made more errors while they were executing them,  $F(1,31) = 34.32$ ,  $MSE = 1.02$ ,  $p < .001$ . No other effects were significant (all  $F_s < 2$ ).

### Text Recall

The minimal and elaborate texts corresponding to each of the four tasks were coded into propositions following the guidelines in Bovair and Kieras (1985). For the minimal and elaborate texts respectively there were 48 and 86 propositions for the VPlus task, 62 and 85 propositions for the TimerProg task, 68 and 78 propositions for the Playback task, and 58 and 99 propositions for the Setup task. All of the propositions from the original minimal texts were included within the elaborate text apart from 7 from the Playback text.

A lenient scoring system was used to score the presence or absence of each proposition. The order of recall did not have to be the same as in the text unless context was needed to disambiguate the proposition being recalled. In all three experiments the proportion of propositions correctly recalled, is used as the measure of recall performance, and the number of steps per task on which participants made an error is used as the measure of errors (so that, in the reported counts of errors, the maximum number of errors is one per step). In Table 2 the means for both of these measures are given for each condition.

The differences in length between elaborate and minimal texts for each task meant recall could be analyzed either using only the propositions common to both texts, or using the proportion of propositions recalled for each text. These analyses yielded very similar results, thus only the results from propositions common to both texts are reported here.

Table 2 shows recall across all four tasks for each of the four conditions. The table indicates that text recall for instructions presented singly was better than for instructions presented in chunks, however this difference was not significant,  $F(1, 31) = 1.27$ ,  $MSE = .013$ ,  $p = .27$ . Similarly, recall of the minimal instructions was better than recall of the elaborate instructions, but the difference was not reliable, ( $F < .5$ ). The difference between the chunking conditions had an effect size of  $\eta^2_p = .20$ .

TABLE 2

### Task Performance at Test

Table 2 also shows means for the task completion time and number of errors per task during test. Once again the results are uncomplicated. When participants were



trained using chunks of instructions they made fewer errors at test,  $F(1, 31) = 4.19$ ,  $MSE = 4.66$ ,  $p < .05$ , and completed the task more quickly at test than when they were trained using single instructions,  $F(1, 31) = 4.75$ ,  $MSE = 2276.67$ ,  $p < .05$ . No other differences were significant ( $F_s < 1$ ). The effect sizes for these differences were  $r_{pb} = .35$  and  $r_{pb} = .36$  respectively.

### Discussion

The main results for the chunking manipulation were as predicted. When instructions were presented in chunks instead of singly task performance at test was faster, and fewer errors were made, indicating that a better situation model had been formed. In contrast at the textbase level chunking made little difference to text recall.

The manipulation of linguistic coherence appeared to have no effect at test, either on overall performance or on the relation between situation model and textbase measures (contrary to our prediction that elaborate texts should allow a crossover dissociation to be detected). This may have been because the alterations to linguistic coherence were too subtle to influence task performance. For example, Carroll (1990) made far more drastic changes: a typical Minimal Manual was less than a quarter of the length of the manual from which it was adapted and with which it was compared.

There was an effect of coherence during training: reading time per syllable was longer for the minimal instructions than the elaborate instructions. This difference could reflect a greater amount of compensatory processing to infer unstated relations in the minimal text. However, it could also reflect the higher number of non-essential propositions in the elaborate text. Readers may have been able to somehow reduce the extent to which this information was processed.

There was no difference between the reading times for instructions that were presented in chunks or singly. Thus it is possible to discount the explanation that chunking instruction steps simply encouraged participants to spend more time reading the instructions which made them more memorable at test. Relative to the single step conditions the chunked conditions did however make more errors and take longer to execute the instructions during training. This difference is unsurprising given that the chunking manipulation deliberately made training more difficult. Nevertheless, it may be seen as a complication for practical purposes.

### Experiment 2A and 2B

Where learning is improved by making training more difficult, it is perhaps inevitable that training time will be increased. However for a manipulation to have any practical worth it is important that this increase is not disproportionate to the gains in task performance at test. In Experiment 1 each step executed within a chunk on average took more than twice as long as when it was executed singly. This experiment aimed to demonstrate this difference could be reduced while maintaining the associated improvement in task performance at test.

Informal observation of the participants in Experiment 1 suggested that a large part of the training execution time was spent attempting to remember the correct button after an error had been made. Table 1 shows that three times as many errors were made in the chunked conditions as the single conditions. Thus by reducing the time between an error and the next step, it should be possible to lessen the overall differences in execution time between chunked and single conditions. Moreover, this could indicate whether the process of resolving an error is in some way crucial to the

improved situation model in the chunked conditions. Therefore in this experiment, directly after a step had been attempted participants were shown how to complete it, irrespective of whether their original attempt had been correct or not.

A second purpose of this experiment is to investigate the role of inference making during the execution of procedures during training, and in particular how inference making relates to our chunking manipulation. Executing the written instructions in Experiment 1 always involved a degree of inference to map from the instructions to the device. Carroll (1990) repeatedly stresses the contribution of inferential processes to the success of minimal instruction, and increasing the necessity for inferences during training is the presumed basis for the McNamara et al. (1996) effects. We obviated the need for any inference making for half the training episodes in this experiment by programming the VCR simulation to illuminate the correct next key at every step during training. This method has direct practical relevance for designers of online help systems, where such exact prompting can readily be implemented. However, our prediction is that this “help” would actually interfere with learning, and with the success of the chunking manipulation.

## Experiment 2A

### Participants

Participants were 30 female and 10 male undergraduate students ranging in age from 17-30, with a mean of 20.5 years. They were paid £4 each or given course credit in exchange for participating.

### Design

Due to the failure to find any effects of coherence in Experiment 1, only the minimal text was used in this experiment. The way in which participants completed the task during training was, however, manipulated, producing a new independent variable. This led to a 2 (number of stages: chunk or single) x 2 (inference: infer or show) design with repeated measures on both variables. The four conditions were labelled chunk-infer, chunk-show, single-infer, and single-show.

### Procedure

The basic train-recall-test cycle of Experiment 1 was adopted. Participants were again allowed to consult a diagram of the Remote Control during the entire experiment, and a task outline during training and performance test.

The fundamental change was that participants were shown the correct buttons to click on, during training. This was done by highlighting the correct button in a yellow colour. In the infer conditions, after reading each step or chunk of instructions, participants were required to carry them out without any assistance as in Experiment 1. However, after a step had been attempted, irrespective of whether it was completed successfully or not the correct button was then highlighted. Where there was more than one button press necessary to complete a step the correct buttons were highlighted in sequence after the first error. If no errors were made then the buttons were highlighted in sequence after the step was completed. In all cases, if a button was highlighted the participants had to click on it before they could start the next step.

In the show conditions the initial attempt by the participants to infer the correct button was omitted. Thus, after reading any instructions participants simply had to click on the buttons that were highlighted in sequence. After they had done this more instructions were then presented.

The practice task was completed twice in the same manner as Experiment 1, once before the first single condition and once before the first chunk condition. The practice procedure always used the infer procedure. Participants were not informed of the inference manipulation. Instead they were simply instructed that “If a button is highlighted that means it is the correct button. Click on any buttons highlighted and then move on to the next stage. If no buttons are highlighted click on whichever button or buttons you think are correct. After you have made your selection the correct button or buttons will be highlighted, click on it/them and then move on to the next stage.”

All other aspects of the procedure were the same as in Experiment 1.

### Experiment 2B

In Experiment 1 and Experiment 2A, participants were explicitly told that they would have to complete the task subsequent to reading the instructions. Given that purpose for reading can differentially affect the development of the textbase and the situation model (e.g. Mills, Diehl, Birkmire & Mou, 1995), and that in real world use of instructions future demands may sometimes be unclear, it is worthwhile to examine the chunking manipulation when the purpose for reading is less clearly stated to the participants. Thus in Experiment 2B participants were not told that they would have to perform the task afterwards.

### Participants

Forty undergraduate students (22 females and 18 males) were each paid £4 to participate in this study. They ranged in age from 19-25 years with a mean of 21 years.

### Procedure

In Experiment 2A, as in Experiment 1, participants were warned at the beginning of the experiment that they would be subsequently required to complete the tasks without the instructions. This information was not mentioned in this experiment. This was the only respect in which it differed from Experiment 2A.

### Experiments 2A and 2B Results

The results from Experiments 2A and 2B were combined to increase the power of the analyses. This was possible because the procedure for both experiments was virtually identical. It was, however, necessary to include experiment as a between participants variable, with inference and number of stages as within participants variables in a 2 x 2 x 2 ANOVA. As in Experiment 1 unless otherwise stated the same analysis was then used for each of the dependent variables. Effect sizes were computed as point biserial correlations.

### Training

Table 3 shows the mean times and errors for each condition during training. When participants were shown which button to click they spent less time reading the instructions than when they had to choose a button,  $F(1, 78) = 17.38$ ,  $MSE = 506.9$ ,  $p$

< .001. The execution times given in Table 3, only refer to the time spent choosing the correct button. (This did not include any time when the button was highlighted. Thus, the execution times from the second phase of the infer condition and all of the show condition were not used.) This execution time data was analyzed using a 2 (Experiment 2A or 2B) x 2 (single-infer or chunk-infer) ANOVA with repeated measures on the second variable. This showed that it took participants longer to execute instructions when they were presented in chunks rather than individually,  $F(1, 78) = 44.50$ ,  $MSE = 2265.03$ ,  $p < .001$ . There were also more errors made during training when instructions were presented in chunks rather than individually,  $F(1, 78) = 53.65$ ,  $MSE = .83$ ,  $p < .001$ , and when participants had to infer a solution, rather than being shown it,  $F(1, 78) = 129.05$ ,  $MSE = 1.08$ ,  $p < .001$ .

There was no effect of experiment in these analyses, and it did not interact with any other variables ( $F_s < 1$ ).

TABLE 3

### Text Recall

Recall and task performance at test is shown in Table 4. Across both experiments a greater proportion of the instructions was recalled when they were presented individually rather than in chunks,  $F(1, 78) = 6.55$ ,  $MSE = .0021$ ,  $p < .05$ , the size of this effect was  $\eta_{pb}^2 = .28$ . The interaction between chunking and inference was marginally significant,  $F(1, 78) = 3.78$ ,  $MSE = .011$ ,  $p = .06$ . Investigation of this interaction showed a simple main effect of inference at the chunk conditions,  $F(1, 78) = 5.31$ ,  $MSE = .014$ ,  $p < .05$ , and of chunking at the show conditions,  $F(1, 78) =$

10.38,  $\underline{MSE} = .0093$ ,  $p < .01$ . The effect sizes were  $r_{pb} = .25$  and  $r_{pb} = .34$  respectively. There were no other significant effects ( $F_s < 2.5$ ).

TABLE 4

#### Task Performance at Test

There were no main effects of the independent variables on the number of errors made at test ( $F_s < 2$ ), however there was a significant interaction between inference and chunking,  $F(1, 78) = 4.18$ ,  $\underline{MSE} = 2.97$ ,  $p < .05$ . Table 4 shows this interaction with fewer errors in the chunk-infer condition than the other three conditions. Simple effects analysis found an effect of chunking at the infer conditions,  $F(1, 78) = 6.61$ ,  $\underline{MSE} = 2.66$ ,  $p < .05$ , and of inference at the chunk conditions,  $F(1, 78) = 5.67$ ,  $\underline{MSE} = 3.10$ ,  $p < .05$ , no other simple effects were significant ( $F_s < 1$ ). The effect sizes for the significant differences were  $r_{pb} = .28$  and  $r_{pb} = .26$  respectively.

Table 4 shows a similar pattern of results for completion time, as the chunk-infer condition completed the task more quickly than the other three conditions. However, this time the only significant main effect or interaction was that when instructions were learnt in chunks not individually, task performance at test was faster,  $F(1, 78) = 6.48$ ,  $\underline{MSE} = 1235.31$ ,  $p < .05$ . Simple effects analysis once again showed an effect of chunking at the infer conditions,  $F(1, 78) = 6.88$ ,  $\underline{MSE} = 1299.95$ ,  $p < .05$ , and of inference at the chunk conditions,  $F(1, 78) = 4.35$ ,  $\underline{MSE} = 1350.14$ ,  $p < .05$ , with no other significant simple effects ( $F_s < 1$ ). The effect sizes for the significant differences were  $r_{pb} = .28$  and  $r_{pb} = .23$  respectively.



### Experiment 2A and 2B Discussion

The task performance results showed that, as in Experiment 1, when participants inferred their own solutions during training, chunks of instructions improved performance more than single instructions. However, when participants were shown the correct buttons to click on, there was no advantage for presenting instructions either singly or in chunks. Training performance that was self-generated was only better than being shown the correct buttons when instructions were presented in chunks. Thus the chunking and inference manipulations were mutually dependent upon each other. The mutual dependence between chunking and inference suggests that some minimal inference making while reading and executing instructions is necessary for the chunk manipulation to have any effect.

The recall results were less clear-cut, because although less was recalled in the chunked conditions than in the single conditions, this difference was not specific to the infer manipulation. Although the chunk-show condition did not produce a better situation model than the single-show condition, the textbase was still worse. One possibility is that the chunk-show condition still encouraged participants to focus on creating a situation model at the expense of the textbase, but the absence of inference making meant this situation model was no better than that formed in the single-show condition. Alternatively, Table 3 shows that the chunk-show condition led to the least time reading the instructions. This may have caused the poor recall performance, rather than the allocation of resources to situation model development instead of textbase development.

During training participants still took longer to execute the instructions in the chunked conditions than in the single conditions, but this difference was considerably reduced relative to Experiment 1.

Experiment as a factor did not interact with any of the comparisons reported, however, by inspection, in Experiment 2A only recall and task performance at test in the single-show condition were similar to the chunk-infer condition. We have no explanation for this anomalous pattern.

In both Experiment 1 and Experiment 2 the improvement in learning caused by chunking instructions during training had a medium effect size in Cohen's (1988) terms. The effect would therefore conventionally be seen as sufficiently large for applied implications, even though its absolute as opposed to statistical size is small.

In all the above experiments the main effect of chunking emerged despite quite large individual variation in task performance. It seems plausible that the effectiveness of chunking as a strategy will be mediated by participants' ability to remember and use the chunks of information. In the absence of any data on participants' working memory capacities, we investigated what role, if any, might be played by their prior self-rated competence, and by their experience with similar devices. We computed correlations between these rating scales and task performance for the different experimental groups and found a mixed pattern of effects. In Experiment 1, competence did not significantly correlate with any index of performance in any experimental group ( $-.20 < r_s < .20$ ), whereas experience correlated moderately with both time and errors in the chunk conditions ( $-.40 < r_{ss} < -.30$ ), and with errors in the single conditions ( $r_s = .30$ ). In Experiment 2 competence correlated moderately and significantly with both errors and time in the chunk-infer and single-show conditions ( $-.28 < r_s (78) < -.23$ , all  $p_s < .05$ ). Experience correlated

moderately and significantly with both errors and time in the chunk-infer and chunk-show conditions ( $-.34 < r_s(N = 80) < -.22$ , all  $p_s < .05$ ). No other correlations approached significance. On balance this pattern of effects provides some (admittedly rather weak) support for the notion that more experienced participants are better placed to benefit from chunking, chiming with the findings of McNamara et al. (1996).

The issues of effect size and individual differences are interesting, especially with regard to the applied implications of this work. But in any case, we feel that the manipulation of chunking used so far, wherein the number and size of chunks was imposed upon participants may be too blunt an instrument for widespread applicability.

Thus, to further develop the practical relevance of chunking as a strategy to improve learning, we sought a less direct manipulation, in which participants were encouraged, rather than forced to chunk.

### Experiment 3

The two preceding experiments presented instruction steps in strict sequence and influenced task performance by controlling the frequency with which participants could interleave reading and acting. However, when participants are allowed to switch freely between device and instructions they do not necessarily work through the instructions in such a strict and simple linear fashion. Rather, learners may choose to consult the instructional text on either side of any particular steps before executing those steps (Gray & Fu, 2001). Our previous experiments do not allow this flexibility, as each instruction step only appeared on the screen once, and was removed before it

was executed (whether presented singly or in a chunk). This constraint, while necessary for experimentation, may conceivably inhibit learning in both chunk and single-step conditions.

Further, as discussed above, individual differences may influence the effectiveness of a chunking manipulation imposed upon participants. In particular, the optimum size of a chunk is likely to vary from person to person (and situation to situation), so that enforcing a rigid chunk size of 3 or 4 steps will not always facilitate performance.

In this experiment we sought to develop learning conditions that allow the natural flexibility of participants' reading strategies, but at the same time encouraged them to chunk steps together. Our approach to this design challenge was informed by recent work showing that interactive performance strategies are very sensitive to the implementation cost of operations (O'Hara & Payne, 1998, 1999; Gray & Boehm-Davis, 2000). According to this line of thought, if the cost of each consultation of written instructions is made higher (even if only by a very small amount), learners will adapt to the structure of the environment by seeking to reduce the number of consultations necessary to perform their task. A recent experiment by Gray and Fu (2001) has demonstrated just this effect in a similar learning context to the current article, acquiring procedures to program a simulated video. Participants who needed to mouse-click on a grayed-out instruction window were more likely to rely on their imperfect memory for procedures than were participants who could consult the instructions by simply shifting their gaze. (But note: Gray and Fu did not study separate acquisition and retention stages, and so could not investigate the learning implications of their manipulation.)

Participants in our experiments presumably have little idea how to achieve the task methods unless they read the instructions in full. The obvious way for participants to reduce the number of consultations of the instructions, in response to a consultation-cost, is to read and remember a chunk of steps during every consultation.

A cost of consulting instructions was implemented by presenting the instructions on a different screen to the device. Participants in the higher-cost condition could choose to display either the instructions or the device on the screen and switched between them by mouse-clicking on a button. The lower-cost condition had both the instructions and the device on the same screen and could glance between the two at will.

### Method

#### Participants

Participants were 40 undergraduate students (26 females and 14 males). Age ranged from 18-31 with a mean of 20.2 years in the high-cost condition, and from 18-41, with a mean of 21.4 years in the low-cost condition. They received course credit in exchange for participating.

#### Stimulus Materials

The simulation needed superficial alteration to enable the instructions for an entire task to be presented at the same time. In the low-cost condition the Instructions panel was enlarged, and the panel representing the TV screen was shrunk slightly to accommodate this change. In the high-cost condition a button labelled “Instructions” replaced the Instructions panel. In both conditions the OK button was removed.

## Design

As in the previous experiments, the order of the tasks was counterbalanced, and as in Experiment 2A and 2B only the minimal text was used. The sole manipulation was the cost of each consultation of the instructions during training, in one condition each consultation had a high-cost, and in the other there was a low-cost to each consultation. This manipulation was between-participants, and they were randomly assigned to an experimental condition with the single constraint that the ratio of males to females was kept the same across both conditions.

## Procedure

The training procedure differed from the previous experiments. After clicking on the Start button the program accepted user input, and all 14 instructional steps were made available, and remained so throughout the training phase. In the low-cost condition this meant they appeared in the Instructions panel in the top right of the screen, and an eye movement was the only cost of consulting the instructions. In the high-cost condition this meant that clicking on the Instructions button replaced the video interface with a separate display containing just the instructions. Each time participants consulted the instructions in the high-cost condition they had to move the mouse to the Instructions button, click on it, then move the mouse to the bottom of the screen, click on another button (labelled “Click here to return to task”), and finally move the mouse back across the screen from the Instructions button to the Remote Control panel. Throughout training the participants were allowed to refer to the instructions as little or as often as desired. The condition participants were in, and thus

the way the instructions were presented during training, remained the same for all four tasks.

The practice task was determined by which condition the participant was in. It was only completed once, and all four steps were presented in the same format as in the training phase.

All other aspects of the procedure were the same as Experiment 1.

### Results

In light of previous moderate correlations between self-ratings and test performance it is important to ensure the experimental groups are reasonably matched for competence and experience. Mean self-rated competence prior to the experiment was 3.4 in the high-cost condition, and 3.25 in the low-cost condition, this difference did not approach statistical significance, ( $t < .5$ ). Frequency of usage of a VCR was also scored on a five point scale with 1 assigned to the lowest frequency. The means were 3.5 in the high-cost condition and 3.3 in the low-cost condition. This difference was not reliable ( $t < 1$ ).

### Training

Mean times and errors made during training are given in Table 5. When instructions were presented on a different screen the total training time was longer than when instructions were presented on the same screen,  $t(38) = 3.19$ ,  $SE = 11.62$ ,  $p < .01$ . There was no reliable difference between the number of errors made in each condition,  $t(38) = 1.65$ .

TABLE 5

During training in the high-cost condition the program also recorded the number of times each participant switched screens between the instructions and the device. This data showed that although there were 14 steps in each task, the mean number of times a participant referred to the instructions was 8.25. Thus, sometimes more than one step must be completed between each referral to the instructions. The data showed that a mean of 10.51 steps per task (75%) were completed as part of a chunk of more than one step (i.e. within a sequence of two or more steps completed before referring back to the instructions). Moreover, a mean of 6.73 steps per task (48%) were completed at least one step after the last reference to the instructions.

The ease of referral to the instructions in the low-cost condition made it difficult to measure the interleaving of reading and acting. However, it is possible to compare the interval between the execution of successive steps in the two conditions. From the data in Table 1 it was computed that the mean execution time when a single step had to be completed with the minimal instructions was 3.40 seconds. This interval was used as a benchmark to estimate the number of steps that had been chunked. Thus, any step that was completed less than 3.40 seconds after the previous step had been finished was deemed to form part of a chunk (because 3.40 seconds does not, by assumption, allow time for reading the instructions for the step before executing it). This interval threshold was undoubtedly conservative, and of course does not incorporate the first step of any chunk. However it was necessary to use a short threshold to minimize the inappropriate inclusion of steps where participants had glanced at the instructions before execution. Table 5 shows that according to this



criterion more steps were completed within the chunking threshold in the high-cost condition than in the low-cost condition,  $t(38) = 3.98$ ,  $SE = .30$ ,  $p < .001$ .

### Text Recall

Table 6 shows the participants in the high-cost condition recalled a somewhat greater proportion of the instructions than those in the low-cost condition. However, this difference was not significant,  $t(38) = .54$ ,  $SE = .025$ ,  $p = .59$ , the effect size was  $r_{pb} = .09$ .

TABLE 6

### Task Performance

Task performance results are presented in Table 6. Participants in the high-cost condition, made fewer errors at test,  $t(38) = 2.04$ ,  $SE = .48$ ,  $p < .05$ , and completed the tasks more quickly,  $t(38) = 2.32$ ,  $SE = 9.71$ ,  $p < .05$ , than participants who had a low cost to consulting the instructions during training. The effect sizes were  $r_{pb} = .31$  and  $r_{pb} = .40$  respectively.

### Discussion

As predicted, when instructions were presented on a different screen to the device it seems that participants were more likely to spontaneously chunk procedure steps than when they were on the same screen. To find more steps executed very quickly during training in the high-cost condition was particularly remarkable given that overall training time was less in the low-cost condition. On this point, it is

striking that the mean training (reading + execution) times for the low-cost condition are very similar to the mean training times for the single-minimal condition in Experiment 1. Similarly, the mean training times for the high-cost condition are comparable with those for the chunk-minimal condition in Experiment 1.

Furthermore, and of primary importance from an applied perspective, learning was more effective when instructions and device were on separate screens than when instructions and device were on the same screen.

The faster training times in the low-cost condition support the findings of Gray and Fu (2001). However, the reversal of this effect at test indicates that for complex multi-step tasks at least, minimizing the cost of referring to the instructions can discourage chunking behaviour, and thus retard longer-term retention of the method.

In contrast with Experiment 2 and with the results of McNamara et al. (1996) and Diehl and Mills (1995) there was no effect of the manipulation on text recall.

## General Discussion

At a general level, the current article contributes to two recent conceptual developments of high practical relevance. First, it develops the Schmidt and Bjork (1992) conception of practice, and in particular its extension to text processing by McNamara et al. (1996). Like these authors, we have demonstrated that increasing the cognitive demands of text comprehension at study can have beneficial effects on long-term retention of useful knowledge. We have shown that this effect can be extended from expository texts to instructions for interactive procedures, and from manipulations of the text itself to manipulations of the way the text is read and used during training. In particular, we have discovered and replicated a learning benefit for

reading chunks of procedural steps as opposed to single steps before executing those steps during training.

McNamara et al. interpret their work in terms of the distinction between textbase and situation model. In the current work, following Diehl and Mills (1995) and others we have suggested that for procedural instructions this distinction is aligned with the dual tasks of text-recall and procedure execution. In support of the distinction we have shown benefits of the chunking manipulation for procedure execution, but no benefits for text recall. However, we have only found weak support for the double-dissociation between situation model and textbase reported by McNamara et al. (1996) and Diehl and Mills (1995). In our experiments chunking typically depressed later text recall, but this effect was only reliable in the show conditions in Experiment 2.

The second theoretical enterprise to which the current work contributes (in Experiment 3) is recent work on problem solving and strategy selection that derives from the “rational” perspective on human cognition (Anderson, 1990). Consider the work of O’Hara and Payne (1998, 1999). In a series of experiments, they manipulated the implementation cost of steps in simple puzzles and simple computer interfaces (for example, by allowing disks of the Tower of Hanoi to be moved by clicking on them or by typing a lengthy command). They reported that increased cost led to more mental lookahead, and thus more efficient performance (in terms of number of moves) and better learning. O’Hara and Payne argued that these results emerge because participants plan sequences of moves to the extent that the mental cost of planning is outweighed by the benefits of more efficient action. In a development of this argument Gray and Boehm-Davis (2000) have shown that very small time-costs on the order of 40-400 milliseconds can exert important pressures on strategy selection.

Small cost manipulations therefore offer a practical tool to interface designers who wish to influence user-behaviour.

In exactly this spirit we have shown how the imposition of a very minor cost on consultation of on-line instructions (a single point-and-click with the mouse) can induce an effective chunking strategy for the study and use of the instructions (Experiment 3).

Although Experiments 1 and 2 demonstrated statistically reliable differences in task performance at test it is unclear whether the particular manipulations we studied could be effectively applied in a practical situation. One problem is that the training regimes in those experiments did not allow for the nonlinear patterns of reading observed by Gray and Fu (2001). This restriction may actually inhibit performance and learning, or it may prove somewhat aversive to learners. A further, related problem, is that the chunk size was fixed and imposed on the learners. A fixed chunk size does not allow for individual differences (e.g. in working memory capacity) that may influence a learner's ability to effectively chunk steps together during study. Thus, Experiments 1 and 2 show that for a population of undergraduates, reading-and-acting in chunks of 3-4 steps is more effective than reading a single step before acting. However, for some populations (the elderly, for example) such large chunks may be detrimental to learning.

Experiment 3 addressed these difficulties by encouraging chunking behaviour as an adaptive response to the cost-benefit structure of the learning situation. Presenting instructions on a different screen to the device instead of the same screen reduced the time to complete the task at test by 20%. Given the ready applicability of this particular manipulation this difference alone is noteworthy, however we feel it probably underestimates the likely benefits in a real situation.

A critical feature of our experimental simulation of the VCR, necessary for controlled experimentation, was that it did not implement any erroneous button presses. That is, any errors simply produced a beep sound and the device did not change its state. But of course, the original VCR would change its state in response to many of the erroneous button presses. Such unwanted effects of errors would often require extensive error recovery procedures to return to the original state, meaning that a relatively small difference in error likelihood (such as those observed with our restricted simulation), might result in very big effects on performance times. Thus, it seems plausible that within any domain where there is a high cost to making an error presenting instructions on a different screen to the device may appreciably improve task performance.

## References

- Anderson, J. R. (1990). The Adaptive Character of Thought. Hillsdale, NJ: Erlbaum.
- Bovair, S., & Kieras, D. E. (1985). A guide to propositional analysis for research on technical prose. In B. K. Britton & J. B. Black (Eds.), Understanding expository text. Hillsdale, NJ: Erlbaum.
- Carroll, J. M. (1990). The Nurnberg Funnel: Designing minimalist instruction for practical computer skill. Cambridge, MA: MIT Press.
- Diehl, V. A., & Mills, C. B. (1995). The effects of interaction with the device described by procedural text on recall, true/false, and task performance. Memory and Cognition, 23(6), 675-688.
- Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds Matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. Journal of Experiment Psychology: Applied, 6(4), 322-335.
- Gray, W. D., & Fu, W.-t. (2001). Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head: Implications of rational analysis for interface design. CHI Letters, 3(1), 112-119.
- Kintsch, W. (1998). Comprehension: A paradigm for cognition. New York: Cambridge University Press.

- McNamara, D. S., Kintsch, E., Songer, N. B., Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. Cognition and Instruction, 14(1), 1-43.
- Mills, C. B., Diehl, V. A., Birkmire, D. P., & Mou, L.-C. (1995). Reading procedural texts: Effects of purpose for reading and predictions of reading comprehension models. Discourse Processes, 20, 79-107.
- O'Hara, K. P., & Payne, S. J. (1998). The effects of operator implementation cost on planfulness of problem solving and learning. Cognitive Psychology, 35, 34-70.
- O'Hara, K. P., & Payne, S. J. (1999). Planning and user interface: The effects of lockout time and error recovery cost. International Journal of Human-Computer Studies, 50, 41-59.
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. Psychological Science, 3, 207-217.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. Psychological Bulletin, 123(2), 162-185.

## Appendix

### TimerProg task minimal instructions

1. Press the OSP button to display the main menu screen.
2. Press number button 1 to select 'timer programming'.
3. Select an empty programme number using number buttons 1 to 6.
4. To select channel press corresponding number button.
5. To record a programme once, daily or weekly press number buttons 1-3.
6. Set the date of the first recording using the number buttons.
7. Set the recording start time using the number buttons.
8. Set the recording stop time using the number buttons.
9. Press the OSP button.
10. Press either of the two TIMER buttons then press the other.
11. Press the OSP button then number button 1.
12. Select the programme number to be cancelled using the number buttons.
13. Press the CANCEL button.
14. Press the OSP button.



## TimerProg task elaborate instructions

### Opening Programme Screen

1. Press the OSP button to display the main menu screen.
2. Press number button 1 to select 'timer programming', this opens the programming screen.
3. There are six different programme numbers. Select an empty programme number using number buttons 1 to 6.

### Setting Recording Channel, Frequency, and Date

4. The asterisks indicate the feature currently selected. Enter the channel to be recorded by pressing the corresponding number button.
5. Select the frequency of recording by pressing number button 1 for 'once', number button 2 for 'daily' and number button 3 for 'weekly'.
6. Set the date in the month on which the first recording is to be made by pressing the corresponding number buttons.

### Setting Recording Time and Transmitting Selection

7. Set the time at which recording starts by pressing the corresponding number buttons.
8. Set the time at which recording stops by pressing the corresponding number buttons.
9. Press the OSP button. The programme setting is now memorised.
10. Press either of the two TIMER buttons then press the other TIMER button.

Cancelling the Timer Programmes

11. Press the OSP button then press number button 1 to select 'timer programming'.
12. Select the programme number to be cancelled by pressing the corresponding number button.
13. Press the CANCEL button. This deletes the information in the selected line.
14. Press the OSP button to exit the programming screen.

Table 1

*Experiment 1 training: Mean times and errors.*

	Reading Time				Executic
	Total		Per syllable		
	M	SD	M	SD	
Chunk-Elaborate	131.33	45.59	0.36	0.12	99.07
Chunk-Minimal	106.66	27.63	0.56	0.15	114.58
Single-Elaborate	138.06	36.26	0.38	0.10	43.76
Single-Minimal	111.59	25.93	0.59	0.17	47.56

Note. Times are given in seconds.

Table 2

*Experiment 1 test: Proportion of propositions recalled, and mean errors and task completion time.*

	Recall		Completion time		Errors	
	M	SD	M	SD	M	SD
Chunk-Elaborate	0.33	0.18	112.08	61.14	3.15	2.34
Chunk-Minimal	0.35	0.17	118.03	53.68	3.21	2.12
Single-Elaborate	0.36	0.16	129.60	58.33	3.78	2.32
Single-Minimal	0.37	0.17	135.99	60.00	4.16	2.59

Note. Times are given in seconds.

Table 3

*Experiments 2A and 2B training: Mean times and errors.*

	Reading time		Execution time		Errors	
	M	SD	M	SD	M	SD
Experiment 2A						
Chunk-Infer	95.64	33.85	69.39	20.31	2.08	1.46
Chunk-Show	73.07	28.46			0.15	0.37
Single-Infer	89.77	28.28	42.60	9.09	0.60	0.84
Single-Show	89.70	21.24			0.00	0.00
Experiment 2B						
Chunk-Infer	95.17	41.61	80.73	59.24	2.13	1.92
Chunk-Show	84.13	33.30			0.15	0.43
Single-Infer	95.07	31.81	38.60	14.15	0.85	0.86
Single-Show	86.33	29.10			0.03	0.16

Note. Times are given in seconds.

Table 4

*Experiments 2A and 2B test: Proportion of propositions recalled, and mean errors and task completion time.*

	Recall		Completion time		Errors	
	M	SD	M	SD	M	SD
Experiment 2A						
Chunk-Infer	0.38	0.17	115.69	44.90	3.88	2.04
Chunk-Show	0.32	0.15	131.94	46.74	4.30	1.91
Single-Infer	0.37	0.18	133.96	45.77	4.40	2.35
Single-Show	0.38	0.15	122.38	58.34	3.90	2.44
Experiment 2B						
Chunk-Infer	0.34	0.17	110.71	40.73	3.33	1.97
Chunk-Show	0.34	0.17	118.73	51.87	4.23	2.64
Single-Infer	0.37	0.16	122.34	47.79	4.13	2.55
Single-Show	0.38	0.17	138.38	56.61	4.38	2.65

Note. Times are given in seconds

Table 5

*Experiment 3 training: Mean times, errors and estimated number of steps chunked, averaged across task.*

	Time						
	Reading		Execution		Total		
	M	SD	M	SD	M	SD	M
High-Cost	95.52	24.61	111.18	16.59	206.70	35.95	1.83
Low-Cost					169.59	37.51	1.70

Note. Times are given in seconds

Table 6

*Experiment 3 test: Proportion of propositions recalled, and mean errors and task completion time averaged across task.*

	Recall		Completion time		Errors	
	M	SD	M	SD	M	SD
High-Cost	0.45	0.12	87.59	26.29	2.54	1.28
Low-Cost	0.43	0.18	110.08	34.55	3.53	1.75

Note. Times are given in seconds